Global Research Alliance N2O chamber methodology guidelines: Flux calculations

Venterea, Rodney; Petersen, Soren; de Klein, Cecile AM; Noble, Alasdair; Rees, RM; Gamble, Joshua; Parkin, Tim; Pedersen, Asger

Published in:
Journal of Environmental Quality

DOI:
10.1002/jeq2.20118

Print publication: 01/09/2020

Document Version
Peer reviewed version

Global Research Alliance N_2O chamber methodology guidelines: Flux calculations

Rodney T. Venterea[1]*, Søren O. Petersen[2], Cecile A.M. De Klein[3], Asger R. Pedersen[4], Alasdair D.L. Noble[5], Robert M. Rees[6], Joshua D. Gamble[7], Timothy B. Parkin[8]

[1]U.S. Department of Agriculture, Agricultural Research Service, Soil and Water Management Research Unit, 1991 Upper Buford Cir., 439 Borlaug Hall, St. Paul, MN 55108

[2]Department of Agroecology and Environment, Faculty of Agricultural Sciences, University of Aarhus, PO Box 50, DK-8830 Tjele, Denmark

[3]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand

[4]Hammel Neurorehabilitation Centre and University Research Clinic, Aarhus University, Voldbyvej 15, 8450 Hammel, Denmark

[5]AgResearch Lincoln, Private Bag 4749, Christchurch 8140, New Zealand

[6]Scotland's Rural College (SRUC), West Mains Road, Edinburgh EH9 3JG, UK

[7]USDA-ARS, Plant Science Research Unit, 1991 Upper Buford Cir., 439 Borlaug Hall, St. Paul, MN 55108

[8]National Laboratory for Agriculture and the Environment, 2110 University Boulevard, Ames, Iowa 50011-3120

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jeq2.20118.

*Author for correspondence: rod.venterea@usda.gov

Abstract

A critical step in determining soil-to-atmosphere nitrous oxide (N_2O) exchange using non-steady state chambers is converting collected gas concentration versus time data to flux values using a flux-calculation (FC) scheme. It is well documented that different FC schemes can produce different flux estimates for a given set of data. Available schemes differ in their theoretical basis, computational requirements and performance in terms of both accuracy and precision. Non-linear schemes tend to increase accuracy compared to linear regression, but also can decrease precision. The chamber bias correction method can be used if soil physical data are available, but this introduces additional sources of error. Here, the essential theoretical and practical aspects of the most commonly used FC schemes are described as a basis for their selection and use. A gold standard approach for application and selection of FC schemes is presented as well as alternative approaches based on availability of soil physical property data and intensity of sample collection during each chamber deployment. Additional criteria for scheme selection are provided in the form of an error analysis tool that quantifies performance with respect to both accuracy and precision based on chamber dimensions and sampling duration, soil properties and analytical measurement precision. Example error analyses are presented for hypothetical conditions illustrating how such analysis can be used to guide FC scheme selection, estimate bias, and inform design of chambers and sampling regimes.

Introduction

A necessary and critical step in determining soil-to-atmosphere N_2O exchange using non-steady state chambers is converting the 'raw' chamber concentration (C_c) versus time (t) data into a flux value for each set of chamber time series data. Non-steady-state chambers rely on the accumulation of gas (in this case, N_2O) within an open-bottom chamber placed on the

soil surface. It is well-documented that placement of the chamber disrupts gas exchange processes, creating a dilemma where the quantity being measured is altered by the measurement process. Accumulation of gas in the chamber suppresses the vertical gas concentration gradient at the soil-atmosphere interface, which suppresses the flux below its 'true' value (f_o) that would have occurred in the absence of a chamber (Healy et al., 1996; Hutchinson and Mosier, 1981). Chamber placement may also promote horizontal movement of soil-gas beneath the chamber which can further suppress the observed flux. These effects create non-linearity in the chamber data which complicate accurate flux determination (Anthony et al., 1995; Livingston et al., 2006; Pedersen et al., 2010).

The 'chamber effect' has motivated development of various non-linear (NL) flux-calculation (FC) schemes to improve the accuracy of flux determination compared to simple linear regression (LR). It is well known that different FC schemes can produce substantially different f_o estimates for a given set of chamber data (Levy et al., 2011). Previous work has shown that available schemes differ not only in their accuracy (ability to estimate the true flux) but also in precision (repeatability). Of particular concern with regard to precision is the sensitivity of flux estimates to random errors that inevitably occur in the measurement of chamber N_2O concentrations (Parkin et al., 2012; Venterea, 2013; Venterea et al., 2009). While NL schemes, including quadratic regression (QR) and HMR, tend to increase accuracy compared to LR, they also tend to be more sensitive to random measurement error. Thus, NL schemes tend to have flux estimates with greater variance than LR-based estimates given the same degree of error in measuring chamber headspace N_2O concentrations. This contrasting performance can result in an 'accuracy versus precision trade-off' and a potential dilemma in selecting an optimum FC scheme. Maximizing accuracy and precision of N_2O flux measurements are both important. Accurate determination of the absolute magnitude of fluxes has become increasingly important in calculating greenhouse gas and nutrient budgets

at plot to global scale. At the same time, precise measurements are needed to support robust statistical evaluation of treatments and management strategies to mitigate emissions, and to improve reporting (Bell et al., 2015). Thus, a comprehensive assessment of FC performance requires consideration of both accuracy and precision and methods for balancing and optimizing both dimensions.

The objective of this study is to assist researchers in optimizing flux determination under specific sets of conditions. The essential aspects of the most commonly used FC schemes are described, including their historical development, theoretical underpinnings, and mathematical structure as a basis for their selection and application. Recommendations for FC scheme selection are provided based on the availability of soil physical property data and the frequency of sample collection during each chamber deployment. In addition, error analysis procedures were developed based on diffusion modeling and Monte Carlo simulation methods. The procedures were codified into a spreadsheet-based error analysis tool that quantifies accuracy and precision for the different FC schemes and provides additional criteria for scheme selection. The tool quantifies mean square error, bias and other performance metrics based on user-supplied inputs including chamber design, deployment period, sampling intensity, soil properties and measurement precision for N_2O concentrations and soil properties. The error analysis can be used to guide FC scheme selection, to estimate bias of resulting flux estimates, and to inform design of chambers and sampling regimes.

Description of flux-calculation schemes

In this section we describe the advantages and disadvantages of most commonly used FC schemes and summarize these in Table 1. While there are many differences among the available FC schemes, there are some basic commonalities in their application. Each scheme is based on an underlying 'model' that is assumed to describe an expected temporal pattern in

chamber N_2O concentrations. A basic distinction between FC schemes is whether the underlying model is 'empirical', that is lacking a theoretical basis, or if it is based in some way on diffusion theory (Table 1). Regardless of this distinction, the general procedure used by all methods is to fit observed chamber data to the underlying model using a regression solver of some type. Best-fit coefficients obtained from the solver are then used to estimate the slope of the line at time 0, or first derivate of the chamber concentration with respect to time ($\left[\frac{dC_c}{dt}\right]_0$), immediately after the chamber is deployed ($t = 0$). Once this quantity is determined, the calculated flux (f_c) can be calculated from

$$f_c = H\left[\frac{dc_c}{dt}\right]_0 \quad \text{Eq. [1]}$$

where H is the 'chamber height', or more precisely the ratio of the internal chamber volume to surface area in contact with the soil. An additional consideration for the use of Eq. [1] is that the units of chamber N_2O concentration (C_c) may need to be converted from mixing ratios (e.g. parts per million or $\mu L\ L^{-1}$) to mass or molar concentrations (e.g. $\mu g\ m^{-3}$ or $\mu mol\ m^{-3}$) using the ideal gas law. Example flux calculations using the LR, QR, rQR and HMR methods described below are included as supplemental information (SI) in Excel format ('Example calculations').

Empirical methods

The most commonly used empirical schemes are linear regression (LR), quadratic regression (QR) and restricted QR (rQR).

Linear regression

Linear regression (LR) uses the simplest of all underlying models:

$$C_c = a_{LR} + b_{LR}t \quad \text{Eq. [2]}$$

where C_c is the chamber N_2O concentration, t is time, and a_{LR} and b_{LR} are the intercept and slope, respectively. The slope (b_{LR}) obtained from least-squares linear regression of C_c versus t is used to represent $\left[\frac{dC_c}{dt}\right]_0$ in Eq. [1] so that f_c is determined from H x b_{LR}. It has been shown numerous times that applying LR to inherently non-linear chamber data, in the absence of measurement error considerations, will tend to underestimate the true flux (f_o) to a greater degree than NL schemes (Conen and Smith, 2000; Hutchinson and Mosier, 1981; Matthias et al., 1978). On the other hand, LR-based flux estimates have been shown to be less sensitive to random errors in measuring chamber N_2O concentrations compared to NL schemes (Venterea et al., 2009). Thus, depending on the extent of measurement error, and other considerations including chamber and sampling design, soil properties, and the magnitude of the flux itself, LR may be a robust option. Error analysis procedures provided here and discussed below can be used to assess the robustness of LR relative to the NL schemes. Use of LR can also be appropriate in the case where the primary choice is a NL scheme, but that scheme 'fails' in fitting a particular chamber data set. In this case LR may be chosen as the secondary method. The rQR and HMR schemes both incorporate this approach, as described below.

Some studies have justified the use of LR based on values of the LR coefficient of determination (r^2) obtained by analysis of individual chamber data sets. When r^2 approaches within some proximity to 1.0, the assumption is that the data are sufficiently 'linear' and therefore that LR-based estimates have minimal bias. While this conclusion is intuitively appealing, it may not be justifiable from a theoretical perspective where non-linearity is expected. For example, Conen and Smith (2000) showed that when LR was applied to theoretical chamber data exhibiting $r^2 > 0.997$, LR still underestimated f_0 by more than 25%. This effect has been shown by others (Healy et al., 1996; Livingston et al., 2006). The error

analysis tool provided here also allows an assessment of the risk of using r^2 as a FC selection criterion.

While it is possible to apply LR with a sampling intensity (n_s) of only two observations per chamber measurement period, these guidelines recommend that $n_s \geq 3$ be used with LR to allow sufficient degrees of freedom for determining confidence intervals and standard errors. (As described below, this is the same reason that $n_s = 4$ is recommended for NL schemes.) Although $n_s = 3$ is not recommended as part of the 'gold standard' approach, researchers may have resource constraints that make this necessary. In this case, the risk of under-estimating the flux with LR can be reduced by deploying increased sampling intensity ($n_s \geq 4$) for subsets of measurements or using the chamber bias correction (CBC) method combined with LR as described below.

Quadratic regression

Quadratic regression (QR) uses a second-order polynomial as the underlying model

$$C_c = a_{QR}t^2 + b_{QR}t + c_{QR} \quad \text{Eq. [3]}$$

where a_{QR}, b_{QR}, and c_{QR} are regression coefficients. Since the first derivative of this expression at t=0 is equal to b_{QR}, f_c can be calculated from H x b_{QR} (Wagner et al., 1997). The QR method can be easily implemented in spreadsheets without using a non-linear regression solver; for example, the multiple regression (LINEST) function in Microsoft Excel can be applied by treating t and t^2 as separate independent variables and C_c as the dependent variable as shown in the provided spreadsheets. When $n_s \geq 4$, LINEST also generates statistical parameters including R^2 and standard errors.

Restricted quadratic regression

The rQR scheme proceeds in the same manner as QR, but also evaluates the second derivative of Eq. [3] which equals $2a_{QR}$. When the regression returns a value of $2a_{QR} > 0$, this indicates 'upward curvature' for the model fit, which is opposite to the expected pattern based on diffusion theory (Parkin et al., 2012; Venterea et al., 2009). In this case, the QR-based flux estimate will be less than the LR estimate, therefore the rQR method reverts to LR when $a_{QR} > 0$. This criterion can be easily applied using the LINEST function as shown in the supplied example calculations. While the rQR scheme does not have a biophysical basis, it can improve FC performance by generating more accurate flux estimates than LR while being less sensitive to measurement error than other NL methods depending on the specific measurement conditions.

Theoretical flux-calculation schemes

Theoretical schemes account for the chamber-placement effect by applying diffusion theory in deriving their underlying models. Due to the complexity of gas diffusion in soil, some simplification of the theory is required in order to arrive at methods that are useful in practice. Varying approaches have been used, none of which can be considered best-suited under all conditions, and each has its own advantages. A commonality among theoretical schemes, as well as QR and rQR, is that the underlying models contain at least three parameters that require fitting to data. Thus, a minimum sampling intensity of $n_s \geq 4$ is recommended when using NL schemes in order to make determination of standard errors and/or confidence intervals of the estimates mathematically feasible. The most widely used theoretical schemes are described below in chronological order of their development.

Hutchinson and Mosier (HM)

The method proposed by Hutchinson and Mosier (1981) was the first theoretical scheme to be widely used. However, the original HM scheme is not recommended here because it does not meet the sampling intensity condition ($n_s \geq 4$) discussed above. Since the updated HMR scheme shares elements with the original HM scheme, those elements are described here. The HM scheme was originally developed to apply to a specific set of conditions. Hutchinson and Mosier (1981) described these conditions as follows: "the zone of N_2O production lies somewhat below the surface and is overlain by a layer of relatively dry, loosely packed soil". Mathematical assumptions consistent with this description allowed the model to be structured as a form of Fick's First Law of diffusion:

$$\left[\frac{dC_c}{dt}\right] = \frac{D_s}{H}\frac{(C_d - C_c)}{d} \qquad \text{Eq. [4]}$$

where C_d is the N_2O concentration at some fixed, but unspecified, depth d in the soil, D_s is the soil gas diffusivity and other terms are as defined above. This model assumes that (i) C_d remains constant over time, (ii) only the soil at or below the depth d is a source of N_2O, and (ii) the driving force for N_2O flux into the chamber is the gradient term $\frac{(C_d - C_c)}{d}$ where d and C_d are constant in time but C_c is changing in time. These assumptions may be questioned, because Fick's First Law is strictly only applicable to 'steady state' conditions where concentrations are not changing or are changing very slowly over time (Rolston and Moldrup, 2002). Thus, these steady-state assumptions might seem to violate its application to 'non-steady state' chambers. An additional mathematical restriction limited its application to n_s=3 where the samples are collected at equally spaced time intervals (n_s=3).

Extended HM (HMR)

Pedersen *et al*. (2010) developed the HMR scheme, which builds on the original HM scheme and expands its applicability and capability. It is contained within an add-on package

to be used with the R statistical program (https://www.r-project.org), and has seen increasing use for estimation of N_2O, as well as CO_2 and CH_4 fluxes (Cowan et al., 2014; Davis et al., 2018; Jones et al., 2016). One major improvement of HMR is that it removes the restriction that $n_s = 3$ equally spaced samples. The HMR package in R provides additional capability including computing 95% confidence as well as being interactive and allowing for user decision making.

The model underlying the extended HM scheme builds upon Eq. [4] as follows

$$\left[\frac{dC_c}{dt}\right] = k(C_d - C_c) - \varphi(C_c - C_0) \qquad \text{Eq. [5]}$$

where $k = \frac{D_s}{d\,H}$ so that the first term on the right side is the same as in Eq. [4]. The second term, $\varphi(C_c - C_0)$, is designed to account for lateral diffusion and/or leakage from an imperfectly sealed chamber, where C_0 is the N_2O concentration in the chamber at t=0 and φ is an unknown parameter. Analogous to the first term, the second term has as its driving force a gradient calculated from the difference between C_c (which is changing over time) and C_0 (which is constant over time) divided by an unknown distance that is embedded in φ. Thus, both terms apply Fick's First Law and can be simplified to result in the same form of Eq. [4] using

$$\left[\frac{dC_c}{dt}\right] = k^*(C_d^* - C_c) \qquad \text{Eq. [6]}$$

but with $k^* = k + \varphi$ and $C_d^* = \frac{k\varphi + C_0}{k^*}$.

The HMR package uses statistical criteria to determine the suitability of the underlying model to each observed data set. If the criteria indicate method 'failure', then either the LR-based flux estimate or a 'no flux' (flux = 0) solution will be recommended depending on the specific cases. In this respect, HMR has some similarity with rQR which also defaults to LR based on the model fit. While the HMR package can be set to run 'automatically' and to select among use of Eq. 6, LR, or the no flux options without user review, the package was designed to

be interactive. It is recommended that users review each data set and utilize diagnostic plots designed to support method selection.

Non-steady state diffusive flux estimator

The non-steady state diffusive flux estimator (NDFE) scheme was developed in part to eliminate the steady state assumption used by HM and HMR and was coauthored by one of the developers of the original HM scheme (Livingston et al., 2005; Livingston et al., 2006). The model underlying NDFE utilizes Fick's Second Law and does not assume N_2O production occurs in a specific soil layer or that a constant soil-gas concentration occurs at a fixed soil depth. Without these assumptions, a more complex partial differential equation was used as the basis for the method, instead of the ordinary differential equations used by HM and HMR. Livingston *et al*. (2006) were able to obtain an analytical solution for the complex partial differential equation which could be implemented as a FC scheme:

$$C_c(t) = C_o + \frac{f_o \tau}{H}\left[\frac{2}{\sqrt{\pi}}\sqrt{t/\tau} + exp\left(\frac{t}{\tau}\right) erfc(\sqrt{t/\tau}) - 1\right] \quad \text{Eq. 7}$$

where the chamber concentration at a given time is a function of the chamber concentration at $t = 0$ (C_o), as well as f_o, H, and τ. Unlike the other FC schemes, f_o itself is a regression coefficient along with $C_o(0)$ and τ. In theory, τ is a function of soil physical properties including bulk density (ρ), water content (θ) and porosity. The NDFE method was initially appealing because it provided a less restrictive theoretical basis for estimating f_o. Later, NDFE was found to have practical limitations and has not been widely used. Its regression solver can return different f_o values for the same data set and can produce flux estimates substantially greater than that determined by other schemes (Kutzbach et al., 2007; Venterea, 2010). In addition, NDFE can underestimate, and in some cases greatly overestimate, f_o when applied to theoretical soil profiles with vertically non-uniform physical properties, or when there is

substantial lateral diffusion beneath the chamber, and appears to be more sensitive to random measurement error than other NL schemes (Venterea, 2013; Venterea and Baker, 2008). It is possible that these issues could be improved by modifying the algorithms in the regression solver, but NDFE is not recommended at the current time. However, the analytical solution (Eq. 7) was an important advance and was used in developing the CBC method described below and for developing accuracy and precision tools including those used here.

Chamber bias correction

The CBC approach is not a stand-alone method but is used in conjunction with a primary FC scheme (LR, QR, rQR or HMR) (Venterea, 2010). It was developed as an alternative to NDFE to compensate for negative biases due to the chamber-placement effect which had been shown by Livingston *et al*. (2005; 2006) and others to be important even with NL schemes. The CBC method avoids the numerical problems of NDFE but is based on the same non-steady state diffusion theory. The method is applied by direct calculation without a regression solver. Example calculations are given in the provided spreadsheets and described in more detail elsewhere (Venterea, 2010; Venterea and Parkin, 2012). The CBC method constrains flux estimates by eliminating the τ term as a regression parameter in Eq. 7 and instead estimates its value based on soil property measurements. The estimated τ value is then used to calculate a correction factor which is applied to the flux estimates obtained by the primary method (LR, QR, rQR or HMR). The soil physical properties required by CBC are commonly measured in N_2O studies, but since these measurements introduce additional sources of error, the advantages of CBC need to be balanced against this error. The error analysis procedures developed here can estimate the effects of this measurement under specific conditions.

Recommendations

Our recommendations are summarized in the form of two decision trees; one for identifying data that fall below flux-magnitude thresholds (Fig. 1) and one for selection of a FC scheme (Fig. 2). Our recommended 'gold standard' approach includes using either of the options in Fig. 1 and the highlighted path in Fig. 2, as described below.

Precision of measurement systems

Determining the precision of the sampling and measurement system used to quantify chamber N_2O concentrations is the starting point for both decision trees (Figs. 1-2) and its usefulness is described in the following sections. Precision here refers to the extent of random error inherent to the measurement system. By definition, random error is expected to equal zero on average across many samples but varies randomly for each individual sample. The degree of random error can be quantified by analysis of replicate samples of a known standard gas or ambient air. These samples should be collected and analyzed in the same way that chamber samples are processed. Parkin et al. (2012) analyzed 35 air samples by gas chromatography and expressed measurement precision as its coefficient of variation (CV_{N2O}) (standard deviation divided by the mean). Parkin et al. (2012) found a CV_{N2O} of 4.4% for N_2O analysis but the level of precision will vary by system. Venterea et al. (2009) found CV_{N2O} values of 1 to 3% for three different gas chromatography systems using N_2O standard gases. Greater CV_{N2O} values are expected using ambient samples, especially samples collected at different times and locations, compared to using standard gases, since individual ambient N_2O concentrations might vary in their true values due to local influences. Thus, the CV_{N2O} determined from ambient samples provides a more conservative (greater) estimate of random error. Frequent determination of CV_{N2O} is also recommended to account for fluctuations in instrument performance.

If the CBC method is to be applied (Fig. 2), it is recommended that analogous methods be used to characterize random error in soil property (SP) measurements, expressed as (CV_{SP}). These values can be obtained in a similar manner as described above by measurement of replicate soil samples. However, CV_{SP} may be considerably greater than CV_{N2O} due to inherently greater spatial variability of ρ and θ compared to the variability of N_2O concentrations in ambient air. Our analysis of replicate soil core samples collected from the upper 0 – 150 mm within single experimental plots in cropped fields yielded CV_{SP} values in the range of 5 to 15% for determination of ρ and θ, respectively. We recommend these values be determined on a site-specific basis, and that treatment-related variation, such as that affected by tillage, be considered.

Data screening methods (Fig. 1)

The gold standard recommends using one of two approaches to identify and process chamber data that do not exceed minimum flux-magnitude criteria following the paths described in Fig. 1. The approaches include either the minimum detectable flux (*MDF*) method, which is available for $n_s = 3$ or 4, or a variance-based filtering (*VBF*) approach that can be applied when $n_s \geq 3$. Example applications of both approaches within Excel spreadsheets are provided as SI. Also, the *VBF* approach has been integrated into the most recent version of HMR (1.0.0).

Minimum detectable flux (*MDF*). The method of Parkin et al. (2012) calculates the *MDF* specific to each FC scheme based on n_s, H, CV_{N2O} and the total duration of the measurement, or chamber deployment period (*DP*). The *MDF* serves as a threshold criterion for handling data sets that fall below the threshold. Because *MDF*s are specific to each FC scheme, the primary FC scheme should first be selected. Data sets that generate flux estimates below the *MDF*-threshold for that scheme can be handled in different ways including flux calculation using LR, assigning some fixed value (e.g. 0 or MDF), or removing

the data set altogether, although the last option can create a discontinuous distribution of data. In many cases, it is desirable to include a flux estimate to allow further data analysis. Error analysis described below shows that LR is increasingly more robust relative to NL schemes at decreasing flux values. Therefore, we recommend a conditional selection strategy where LR is applied to data sets that yield an initial flux value below the *MDF*, while the primary scheme (which may be LR or an NL scheme) is used for data sets that exceed the threshold. One current limitation of the Parkin et al (2012) analysis is that statistical parameters required for its application were only determined for n_s = 3 and 4 for LR, QR and rQR, and only for n_s = 4 for HMR (although in theory the analysis could be extended to other n_s values).

Variance based filtering (*VBF*). While the *MDF* approach is applied following flux calculation, the *VBF* approach applies criteria prior to flux calculation and can be applied when $n_s \geq 3$. This approach identifies data sets having variation in N_2O chamber concentrations that are clearly in excess of what is expected due to natural variation in ambient N_2O concentrations and passes those data sets along for flux calculation using the primary scheme. Data sets having a variance which does not exceed this criterion can (as with the *MDF* method) be subjected to flux calculation using LR, assigned a fixed value, or excluded altogether. For n_s = 3 or 4, *VBF* can be applied in conjunction with the *MDF* method as shown in Fig. 1.To apply the *VBF* approach, the level of sample variance expected in the absence of a flux (σ_{air}^2) is first determined by analysis of replicate samples of ambient air, as described above. Here, σ_{air}^2 is equivalent to $[CV_{N2O} * \bar{X}]^2$ where $\bar{X}$ is the mean ambient value. Data sets with variances that are not significantly larger than σ_{air}^2 may be identified by testing the null hypothesis H_0: $\sigma_{sample}^2 = \sigma_{air}^2$ against the one-sided alternative hypothesis H_A: $\sigma_{sample}^2 > \sigma_{air}^2$ for each data set. The null hypothesis is rejected when $\frac{s^2}{\sigma_{air}^2} > \chi_{cr}^2/(n_s - 1)$, where s^2 is an estimate of σ_{sample}^2 calculated from measured N_2O concentrations C_0, C_1,

C_2… at time-points t_0, t_1, t_2… after deployment, and χ^2_{cr} is the $1-\alpha$ quantile of the χ^2 distribution with $n_s - 1$ degrees of freedom. At a 5% significance level (α=0.05), the value of the critical ratio $\chi^2_{cr}/(n_s - 1)$ is provided in Table 2 for a range of sample sizes. Example calculations are provided.

Selection of a FC scheme

Our recommended approach for selecting a FC scheme is illustrated in Fig. 2, with the gold standard approach indicated by the red path. The recommendations depend on two main factors, the availability of soil data to support CBC, and the sampling intensity (n_s) or the number of samples collected during each chamber deployment. In addition, as shown in Fig. 2, the use of error analysis to provide supplementary selection criteria is recommended at two possible points in the decision tree: (i) prior to conducting measurements to support the design of chambers and/or sampling protocols; and (ii) if $n_s \geq 4$, to support selection of a specific FC scheme. Details regarding the error analysis and example applications are provided in the following section.

Following preliminary error analysis, the first decision point relates to the availability of soil data to support the CBC method. The gold standard recommends collection of the required soil data. The CBC method has the potential to substantially improve and stabilize FC performance. Robust application of the CBC method utilizes frequent measurement of soil physical properties, including temperature, ρ and θ, ideally with measurements coinciding with every flux sampling date. The frequency of soil sampling should be guided by the expected temporal and spatial variability of soil properties. The sensitivity of the CBC method performance to the precision of the soil measurements can be assessed using error analysis (see example below). Additional analysis (SI) showed that soil data collected from the upper 0-50 mm to 0-150 mm of the soil profile provide the most robust CBC estimates.

The next decision point is based on n_s. As described above, NL schemes are only recommended with $n_s = 4$ or greater to provide enough degrees of freedom to calculate confidence intervals and standard errors. Thus, using $n_s \geq 4$ allows selection from all available schemes, while $n_s = 3$ restricts the selection only to LR. While $n_s = 2$ has been used in past studies, typically motivated by a trade-off against other sources of variability (spatial or temporal), this practice does not allow for analysis of significance or quality of flux estimates, and therefore it is not considered further. The gold standard recommends collection of soil data with $n_s \geq 4$. Thus, alternatives to the gold standard include (a) collection of soil data with $n_s = 3$, which supports LR supplemented with CBC (LR-CBC), (b) $n_s \geq 4$ without soil data, which supports LR, QR, rQR or HMR, or (c) $n_s = 3$ without soil data, which only supports LR. The latter option is least robust. For this case it is recommended that increased sampling intensity ($n \geq 4$) be used on a subset of chamber locations, preferably during each sampling event (Charteris et al., 2020). Data collected from these chambers can be used to calculate fluxes using a NL scheme and compared to fluxes calculated from the same chambers using LR with $n_s = 3$.

Error analysis

The objective of the error analysis is to quantify the inherent trade-off between accuracy and precision among FC schemes that is described in the above sections. The analysis determines statistical 'performance' metrics that incorporate both accuracy and precision based on user-supplied inputs that represent expected conditions for a given experiment. Using a combination of diffusion modeling and Monte Carlo simulation, each set of conditions is replicated over a large number of trials (10,000) to account for the effects of random measurement error. The aggregated results across trials may indicate that one of the FC schemes performs best relative to the others, although in some cases a clear 'winner' may

not emerge. Alternatively, the analysis can be used prior to design of chambers and sampling regimes so that *H* and/or *DP* can be adjusted to optimize FC performance.

The analysis builds on diffusion modeling-based methods to quantify accuracy (Livingston et al., 2006; Venterea, 2013) combined with Monte Carlo simulation methods to quantify precision (Parkin et al., 2012; Venterea et al., 2009). The analysis consists of three steps (Fig. S1, see SI for additional details): Step 1 uses diffusion modeling to generate a set of theoretical chamber time series based on user-supplied inputs including the true flux (f_o), n_s, H, DP, ρ and θ. The generated data represent the theoretical chamber N_2O concentrations versus time based on the diffusion model assumptions, and under the assumption that chamber measurements are made with 100% accuracy. Step 2 uses Monte Carlo methods to apply random measurement error to the data from step 1 based on user supplied CV_{N2O}, resulting in 10,000 'error-adjusted' chamber data sets for each data set generated in step 1. A corresponding set of 10,000 'error-adjusted' soil property values are also generated based on user supplied CV_{SP}. In step 3, each of the 10,000 error-adjusted data generated in step 2 are analyzed by each FC scheme to produce distributions of 10,000 calculated flux (f_c) values which are then used to calculate the mean square error (*MSE*) given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f_{c_i} - f_o\right)^2 \quad \text{Eq. 8.}$$

with units [μg N m^{-2} h^{-1}]^2 where n = 10,000. The *MSE* can also be expressed as the sum of the bias squared and the variance, i.e.

$$MSE = \left[\frac{1}{n}\sum_{i=1}^{n}(f_{c_i} - f_o)\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left(f_{c_i} - \bar{f}_c\right)^2 \quad \text{Eq. 9}$$

where $\overline{f_c}$ is the mean calculated flux, the term inside brackets, i.e., $\frac{1}{n}\sum_{i=1}^{n}(f_{c_i} - f_o)$, is the mean error, or bias, which is inversely related to accuracy, and the second term, $\frac{1}{n}\sum_{i=1}^{n}\left(f_{c_i} - \bar{f_c}\right)^2$,

is the variance, which is inversely related to precision (DeGroot, 1986). Thus, considering the accuracy-precision trade off inherent to FC schemes, minimizing *MSE* can be considered an appropriate criterion for scheme selection because it expresses performance with respect to both accuracy and precision (Parkin and Venterea, 2010). Additional metrics that reflect both accuracy and precision are also computed, including probabilities of estimating f_o within user-specified tolerance limits and probabilities of over- and under-estimating f_o by user-specified limits, as well as other traditional model performance metrics including bias, root-mean square error, and mean absolute error.

After applying the above analysis to a wide range of hypothetical conditions, we concluded that FC scheme performance cannot be broadly generalized. For this reason, we recommend error analysis be conducted on a case-by-case basis. To facilitate its use, the error analysis procedures were codified into an Excel spreadsheet-based tool which is provided as SI. Example applications and limitations of the analysis are described below.

Implementation of error analysis. The analysis requires user input of a set of 'key factors' which include H, DP, n_s, CV_{N2O}, CV_{SP}, ρ, θ and f_o, as well as soil temperature and clay content. Five of these factors can be considered relatively fixed (H, DP, n_s, CV_{N2O}, CV_{SP}) while ρ, θ and f_o are expected to vary during any measurement period, especially θ and f_o. Thus, the analysis should examine a range of ρ, θ and f_o values expected, or already observed, during the experiment, combined with the other factors (H, DP, n_s, CV_{N2O}, CV_{SP}). The error analysis is *not* designed to be applied to each individual measurement over the

course of an experiment, but instead to assist in selection of a primary FC scheme which is then applied to the observed data. In contrast, the CBC method does utilize each individual soil property measurement. The error analysis can be used to assess the potential performance of the CBC method, as shown in the examples below. As for the CBC method, expected soil property data from the upper 0-50 mm to 0-150 mm of the soil profile should be used for the error analysis. Soil temperature and clay content are also required inputs but have smaller effects and can be approximated within $\pm 5^{o}$C and $\pm$10% of expected values, respectively, without affecting the overall results.

Example error analysis applications. Of the key factors, only H, DP and n_s can be considered under complete control of the researcher. The sensitivity of FC performance to these factors can be assessed prior to design of chambers or sampling protocols using knowledge of CV_{N2O} and CV_{SP} and by approximating expected ranges of the other factors (ρ, θ and f_o). Impacts of n_s can be assessed using separate analysis spreadsheets that are supplied for n_s = 3, 4 and 5 while impacts of the other factors can be examined within a single spreadsheet. As an example, performance of the FC schemes with and without CBC were compared at two levels of CV_{N2O} and different values of H (Fig. 3) and DP (Fig. 4). In these examples, other factors are held constant (as specified in the figure captions). Performance is assessed using MSE and the probability of f_c falling within $\pm$15% of f_o (Pr_{15}). These types of plots can be readily generated for a given set of key factors using the spreadsheet tool. The output can be used to inform design of H or DP and/or guide selection of a FC scheme that will minimize MSE and maximize Pr_{15}. For this example, in the absence of CBC, selection of H = 0.20 m together with HMR or rQR would achieve Pr_{15} = 94% or 79% for CV_{N2O} = 1 and 3%, respectively. If CBC is applied with CV_{SP} = 10%, and assuming DP = 1.0 h and CV_{N2O} = 3%, selection of H = 0.25 m together with LR-CBC would in theory achieve Pr_{15} = 95% (Fig. 3b) for these conditions.

The sensitivity of FC performance shown in Figs. 3-4 is the basis for our recommendation to use error analysis to support measurement design and/or FC method selection. However, the specific results for the above example should not be generalized to other conditions, because performance is sensitive to f_o, ρ and θ which commonly will vary widely during any measurement campaign. These effects can complicate selection of the optimum FC scheme, but error analysis can provide guidance in the form of sensitivity analysis. For example, the sensitivity of *MSE* to f_o and θ over the range of their expected values for an assumed set of conditions is shown in Fig. 5. For this example, in the absence of CBC, selection of HMR or rQR would minimize *MSE* and maximize Pr_{15} (not shown), except at lower values of f_o (Fig. 5a) and greater values of θ (Fig. 5b). The trend for LR to perform better relative to the NL schemes as f_o decreases is the basis for our above recommendation to substitute LR for the primary scheme using either *MDF* or *VBF* criteria (Fig. 1). This approach will constrain errors at lower f_o values, where performance of the NL schemes degrades relative to LR, and at higher f_o values where LR can substantially underperform depending on conditions. For the example shown in Fig. 5, selection of LR-CBC would minimize *MSE* across the entire range of f_o and θ. This favorable outcome will not always occur when CBC is applied, because outcomes depend on the unique combination of key factors. But the potential for improved and more consistent performance over ranging values of f_o and θ, is the basis for our recommendation to use CBC when possible.

The error analysis also assesses the risk of using the LR coefficient of determination (r^2) as a FC selection criterion for any individual chamber data set. For a given set of key factors, the average bias of LR-based flux estimates relative to NL-based flux estimates is computed from

$$b_{LR} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f_{c_i}^{LR} - f_{c_i}^{NL}}{f_{c_i}^{NL}}\right) \quad \text{Eq.}$$

where $f_{c_i}^{LR}$ and $f_{c_i}^{NL}$ are the LR and NL-based flux estimates, respectively, for each corresponding set of error-adjusted data, and *n* is the number of data sets yielding r^2 values within a specified range. For example, a b_{LR} value of -25% indicates that the LR-based estimate was on average 25% less than the NL estimate over a given range of r^2 values. For the example conditions (Fig. 6), r^2 values ≥ 0.99 were required to achieve b_{LR} > -10%, while b_{LR} for data sets with r^2 in the range of 0.94 to 0.99 ranged from -45% to -10% relative to HMR or rQR.

Error analysis limitations and other considerations

A limitation of the error analysis is that its first step, generation of 'error-free' chamber data, necessarily requires assumptions regarding the production and diffusion of N_2O in the soil and its accumulation in the chamber. The default diffusion model incorporated into the spreadsheet is based on the non-steady state theory of Livingston et al. (2006) as expressed in Eq. 7. The spreadsheet also has an option to analyze chamber data generated from alternative diffusion models. This option requires separate model implementation followed by importing the model output into the spreadsheet, whereas the default Livingston et al. (2006) model is implemented within the spreadsheet.

There are at least two limitations of the Livingston et al. (2006) model that can impact the error analysis. The first is the assumption that there is no horizontal diffusion of N_2O in the soil beneath the chamber impacting N_2O accumulation in the chamber. Therefore, the error analysis results are most robust when chamber bases or anchors are inserted to a depth sufficient to reduce horizontal diffusion to negligible levels. Readers are directed to other sources for relevant criteria to guide proper chamber insertion depth (Clough et al.,

2020; Hutchinson and Livingston, 2001; Hutchinson and Livingston, 2002; Parkin and Venterea, 2010). Here, the effects of chamber insertion depth were evaluated using a less restrictive diffusion model (Venterea, 2013) that explicitly accounts for effects of chamber insertion depth on horizontal diffusion to generate theoretical chamber data. Results of this evaluation (described in SI) showed that the performance of HMR, which is designed to account for horizontal diffusion, can be underestimated relative to performance of the other FC schemes, when chamber insertion depth is restricted to < 80 mm especially under drier soil conditions and when $CV_{N2O} \leq 5\%$. While insertion depth < 80 mm is not recommended, it may be unavoidable for a variety of reasons, for example in rocky or hard soils. Thus, under these conditions, and/or when chamber depth does not meet other criteria (Hutchinson and Livingston, 2001; Hutchinson and Livingston, 2002), error analysis results for HMR should be used with caution, and we recommend that selection of HMR should be given additional consideration.

The second limitation of the Livingston et al. (2006) model is its assumption that soil physical properties are vertically uniform. This assumption is also made by the underlying model of the HM and HMR methods. This assumption can impact error FC accuracy when applied to soils that are not uniform (Venterea, 2013; Venterea and Baker, 2008). To evaluate this assumption, we also used the model of Venterea (2013), which can account for non-uniformity, to simulate three soil profiles having large vertical gradients in ρ and/or θ. The results (SI) showed that when soil data from the upper 0-50 mm to 0-150 mm depth intervals were used as inputs, the Livingston et al. (2006) model generated chamber data similar to the less restrictive model of Venterea (2013). These results are the basis for our recommendations to use soil data from these depth intervals for both the error analysis and when using the CBC method.

Another limitation of the error analysis is that evaluating HMR requires additional effort. For each set of conditions and associated set of 10,000 Monte Carlo simulations, the HMR-R package needs to be implemented separately, followed by importing the HMR flux output back into the spreadsheet. In contrast, the calculations are made instantly within the spreadsheet for LR, QR and rQR. Also, the error analysis may or may not reflect the actual performance of HMR, depending on how the HMR-R package is applied in practice. The error analysis used the '*FollowHMR*' command whereas HMR's interactive capabilities allow the user to over-ride the HMR-recommended FC method (i.e., HMR, LR, or no-flux) based on review of each individual dataset, which would be prohibitively time-consuming to apply to 10,000 simulations. Due to the tendency of HMR to sometimes return extraneously high flux values, the error analysis does allow for automated filtering of HMR-calculated f_c values that exceed LR-based values by a specified factor and replacement of those values with the LR-based values prior to calculating performance metrics (a factor of 10 is the default value and was used for the reported results). For these reasons, the error analysis spreadsheet is less practical for evaluating HMR performance than for LR, QR and rQR, although it may be useful as a research tool with respect to HMR performance.

The analysis tool may also facilitate investigation of a variety of research questions related to optimizing FC methods. As described above, the tool allows analysis of theoretical chamber data generated by any model as the starting point for the analysis. The spreadsheet can also analyze the performance of any FC method other than HMR that requires an external regression solver, for example, updated versions of HMR or NDFE or yet-to-be developed methods. The spreadsheet is also designed to quantify the sensitivity of FC performance to *systematic* measurement error, in addition to random error, by entering parameters that characterize the degree of over- or under-estimation of N_2O concentrations and/or soil properties used in the CBC calculations. To facilitate further investigation, we provide as SI

results of additional analyses examining a range of conditions based on theoretical chamber data simulated using both the Livingston et al (2006) and Venterea (2013) models.

Flux calculation performance is sensitive to multiple variables that can range widely in magnitude under real field application and interact to affect error analysis results. It is not possible to evaluate the full universe of potential combinations of these variables. This reality precludes broad generalization regarding selection of optimal calculation methods or measurement regimes and is the basis for our general recommendation to conduct error analysis on a site-specific basis using all available information in order to increase confidence in the resulting data.

Acknowledgements

We are thankful for financial support for CdK and for publishing costs from the New Zealand Government in support of the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. SOP received support from the Ministry of Environment and Food of Denmark. RTV received support from the U.S. Department of Agriculture. The authors thank an anonymous reviewer for very helpful comments.

References

Anthony W., Hutchinson G., Livingston G. (1995) Chamber measurement of soil-atmosphere gas exchange: Linear vs. diffusion-based flux models. Soil Science Society of America Journal 59:1308-1310.

Bell M., Hinton N., Cloy J., Topp C., Rees R., Cardenas L., Scott T., Webster C., Ashton R., Whitmore A. (2015) Nitrous oxide emissions from fertilised UK arable soils: fluxes, emission factors and mitigation. Agriculture, Ecosystems & Environment 212:134-147.

Charteris A.F., Chadwick D.R., Thorman R.E., Vallejo A., de Klein C.A.M., Cardenas L.M. (2020) Global Research Alliance N2O chamber methodology guidelines:

Recommendations for deployment and accounting for sources of variability. Journal of Environmental Quality (this issue).

Clough T.J., Rochette P., Thomas S., Pihlatie M., Christiansen J., Thorman R.E. (2020) Global Research Alliance N2O chamber methodology guidelines: Design considerations. Journal of Environmental Quality (this issue).

Conen F., Smith K. (2000) An explanation of linear increases in gas concentration under closed chambers used to measure gas exchange between soil and the atmosphere. European Journal of Soil Science 51:111-117.

Cowan N., Famulari D., Levy P., Anderson M., Bell M., Rees R., Reay D., Skiba U. (2014) An improved method for measuring soil N2O fluxes using a quantum cascade laser with a dynamic chamber. European journal of soil science 65:643-652.

Davis M.P., Groh T.A., Parkin T.B., Williams R.J., Isenhart T.M., Hofmockel K.S. (2018) Portable automation of static chamber sample collection for quantifying soil gas flux. Journal of environmental quality 47:270-275.

DeGroot M.H. (1986) Probability and Statistics. Addison-Wesley Pub. Co., Reading, MA Addison-Wesley

Healy R.W., Striegl R.G., Russell T.F., Hutchinson G.L., Livingston G.P. (1996) Numerical evaluation of static-chamber measurements of soil—atmosphere gas exchange: Identification of physical processes. Soil Science Society of America Journal 60:740-747.

Hutchinson G., Livingston G. (2001) Vents and seals in non-steady-state chambers used for measuring gas exchange between soil and the atmosphere. European Journal of Soil Science 52:675-682.

Hutchinson G., Livingston G. (2002), in: J. Dane and G. Topp (Eds.), Methods of soil analysis. Part 4. , Amaerican Society of Agronomy, Madison, WI. pp. 1159–1182.

Hutchinson G., Mosier A. (1981) Improved soil cover method for field measurement of nitrous oxide fluxes 1. Soil Science Society of America Journal 45:311-316.

Jones S.P., Diem T., Quispe L.P.H., Cahuana A.J., Reay D.S., Meir P., Teh Y.A. (2016) Drivers of atmospheric methane uptake by montane forest soils in the southern Peruvian Andes. Biogeosciences 13:4151-4165.

Kutzbach L., Schneider J., Sachs T., Giebels M., Nykänen H., Shurpali N., Martikainen P., Alm J., Wilmking M. (2007) CO 2 flux determination by closed-chamber methods can be seriously biased by inappropriate application of linear regression. Biogeosciences 4:1005-1025.

Levy P., Gray A., Leeson S., Gaiawyn J., Kelly M., Cooper M., Dinsmore K., Jones S., Sheppard L. (2011) Quantification of uncertainty in trace gas fluxes measured by the static chamber method. European Journal of Soil Science 62:811-821.

Livingston G.P., Hutchinson G.L., Spartalian K. (2005) Diffusion theory improves chamber-based measurements of trace gas emissions. Geophysical Research Letters 32.

Livingston G.P., Hutchinson G.L., Spartalian K. (2006) Trace gas emission in chambers. Soil Science Society of America Journal 70:1459-1469.

Matthias A.D., Yarger D.N., Weinbeck R.S. (1978) A numerical evaluation of chamber methods for determining gas fluxes. Geophysical Research Letters 5:765-768.

Parkin T., Venterea R. (2010) Sampling Protocols. Chapter 3. Chamber-Based Trace Gas Flux Measurements. IN Sampling Protocols. RF Follett, editor, to.

Parkin T.B., Venterea R.T., Hargreaves S.K. (2012) Calculating the Detection Limits of Chamber-based Soil Greenhouse Gas Flux Measurements. Journal of Environmental Quality 41:705-715. DOI: 10.2134/jeq2011.0394.

Pedersen A.R., Petersen S.O., Schelde K. (2010) A comprehensive approach to soil-atmosphere trace-gas flux estimation with static chambers. European Journal of Soil Science 61:888-902.

Rolston D., Moldrup P. (2002) Methods of soil analysis. Part 4: Physical methods. Gas diffusivity:1113-1139.

Venterea R.T. (2010) Simplified Method for Quantifying Theoretical Underestimation of Chamber-Based Trace Gas Fluxes. Journal of Environmental Quality 39:126-135. DOI: 10.2134/jeq2009.0231.

Venterea R.T. (2013) Theoretical Comparison of Advanced Methods for Calculating Nitrous Oxide Fluxes using Non-steady State Chambers. Soil Science Society of America Journal 77:709-720. DOI: 10.2136/sssaj2013.01.0010.

Venterea R.T., Baker J.M. (2008) Effects of soil physical nonuniformity on chamber-based gas flux estimates. Soil Science Society of America Journal 72:1410-1417. DOI: 10.2136/sssaj2008.0019.

Venterea R.T., Parkin T.B. (2012) Quantifying Biases in Non-Steady-State Chamber Measurements of Soil-Atmosphere Gas Exchange. Managing Agricultural Greenhouse Gases: Coordinated Agricultural Research through Gracenet to Address Our Changing Climate:327-343. DOI: 10.1016/B978-0-12-386897-8.00019-X.

Venterea R.T., Spokas K.A., Baker J.M. (2009) Accuracy and Precision Analysis of Chamber-Based Nitrous Oxide Gas Flux Estimates. Soil Science Society of America Journal 73:1087-1093. DOI: 10.2136/sssaj2008.0307.

Wagner S.W., Reicosky D.C., Alessi R.S. (1997) Regression Models for Calculating Gas Fluxes Measured with a Closed Chamber. Agronomy Journal 89:279-284. DOI: 10.2134/agronj1997.00021962008900020021x.

Table 1. Summary of key advantages and disadvantages of empirical and theoretical N_2O flux calculation schemes.

Scheme	Advantages	Disadvantages
Empirical schemes		
Linear regression (LR)	-Least sensitive to measurement error -Computationally simple, applicable with $n_s = 3$	-No theoretical basis -Most negatively biased method on average
Quadratic regression (QR)	-Less negatively biased than LR on average -Moderately sensitive to measurement error -Can be implemented in spreadsheet format	-No theoretical basis -Underestimates flux compared to LR with 'upward' curvature -More negatively biased than HMR on average
Restricted QR (rQR)	-Similar to QR but reverts to LR for 'upward' curvature -Moderately sensitive to measurement error -Can be implemented in spreadsheet format	-No theoretical basis -More negatively biased than HMR on average
Theoretical schemes[†]		
Extended Hutchinson and Mosier (HMR)	-Based on same theory as HM method[†] plus additional consideration of horizontal diffusion beneath chambers -Available as part of interactive R package -Provides confidence intervals and standard errors	-More sensitive to random measurement error (less precise) than LR, QR and rQR, especially at lower flux values -Can return unexpectedly large fluxes relative to other methods

Chamber bias correction (CBC)	-Based on same theory as NDFE† but avoids its pitfalls -Compensates for residual chamber effects -Can be combined with LR, QR, rQR or HMR -Can be implemented in spreadsheet format	-Requires collection of soil data -Soil data introduces additional source of errors

†The original Hutchinson and Mosier (HM) and non-steady state diffusive flux estimator (NDFE) schemes are also theoretically based and are described in the text. These schemes are not included in the table since they are not recommended for reasons given in the text.

Table 2. Critical ratios, $\chi^2_{cr}/(n_s - 1)$, at varying sample size (n_s) used as threshold criteria for the variance-based filtering method (*VBF*). As described in text and Fig. 1, when the ratio between variance of N_2O concentrations in chamber headspace samples and variance of N_2O concentrations in ambient air does not exceed the critical ratio, non-linear FC schemes should not be applied. Example calculations are provided.

$Sample\ size\ (n_s)$	3	4	5	10	25	50	100	$\propto$
$\chi^2_{cr}/(n_s - 1)$	3.00	2.60	2.37	1.88	1.52	1.35	1.24	1

Figure captions

Figure 1. Decision tree summarizing recommended practices for identifying and processing chamber data that do not exceed minimum flux-magnitude criteria. When the number of samples (n_s) = 3 or 4, either the MDF or VBF method is recommended. When n_s > 4, the VBF option is recommended. Details are provided in text and example calculations are provided in supplementary spreadsheet.

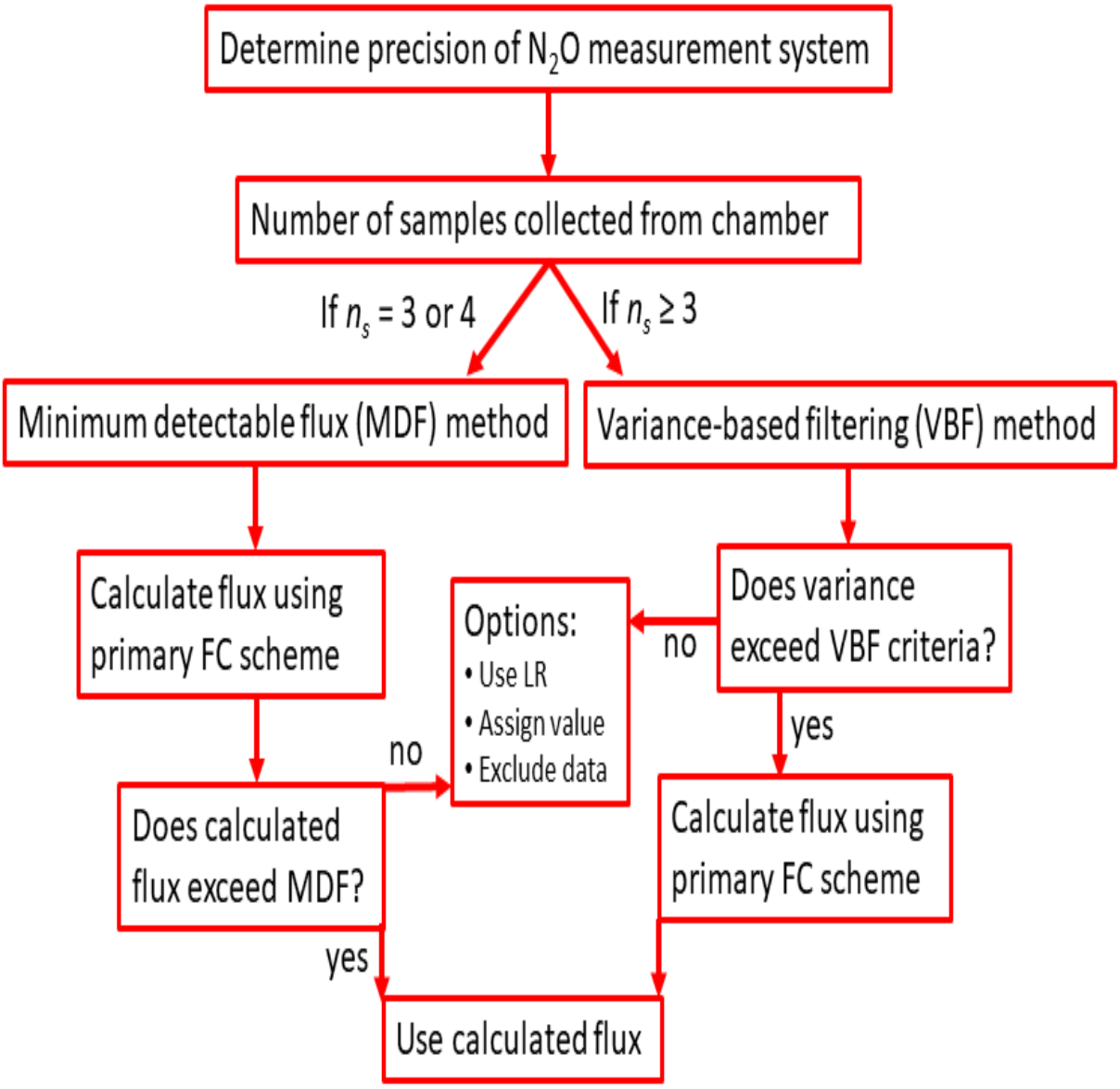

Figure 2. Decision tree summarizing recommended practices for selecting flux-calculation schemes. The red-highlighted path is the gold standard. Details are provided in the text.

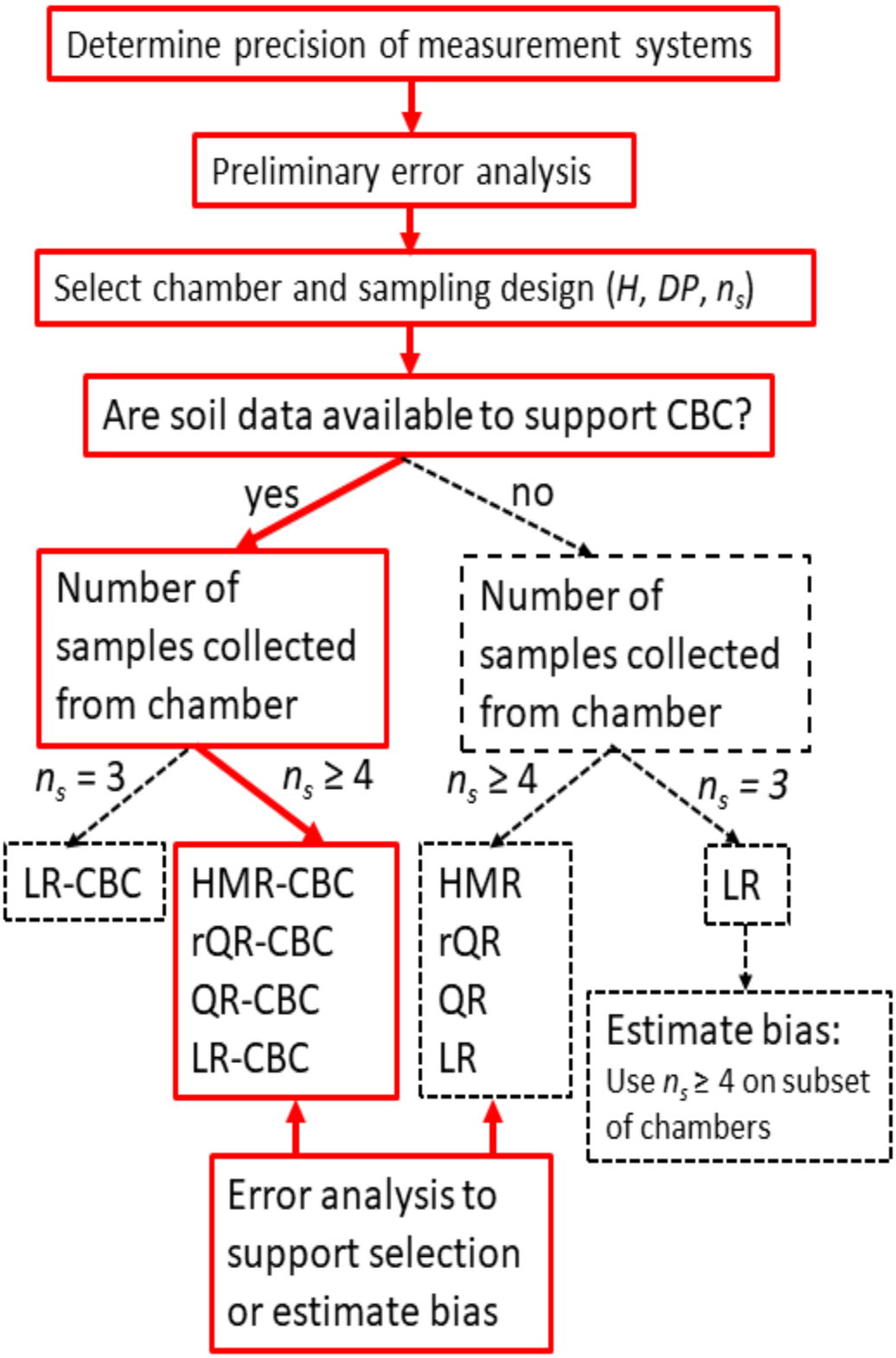

Figure 3. Effects of chamber height (H) and N_2O measurement precision (CV_{N2O}) on (a) mean square error (MSE) and (b) probability (Pr_{15}) that the calculated flux (f_c) is within ±15% of the true flux (f_o) for fluxes calculated using linear regression (LR), quadratic regression (QR), restricted quadratic regression (rQR) and HMR with and without chamber bias correction (CBC). The analysis assumed f_o = 100 μg N m^{-2} h^{-1}, chamber deployment period (DP) = 1.0 h, soil property measurement precision (CV_{SP}) = 10% and soil water content, bulk density, clay content and temperature of 0.30 g H_2O g^{-1}, 1.0 g cm^{-3}, 22% and 20°C, respectively. Each MSE and Pr_{15} value was determined from a distribution of fluxes calculated from 10,000 Monte Carlo simulated chamber data sets.

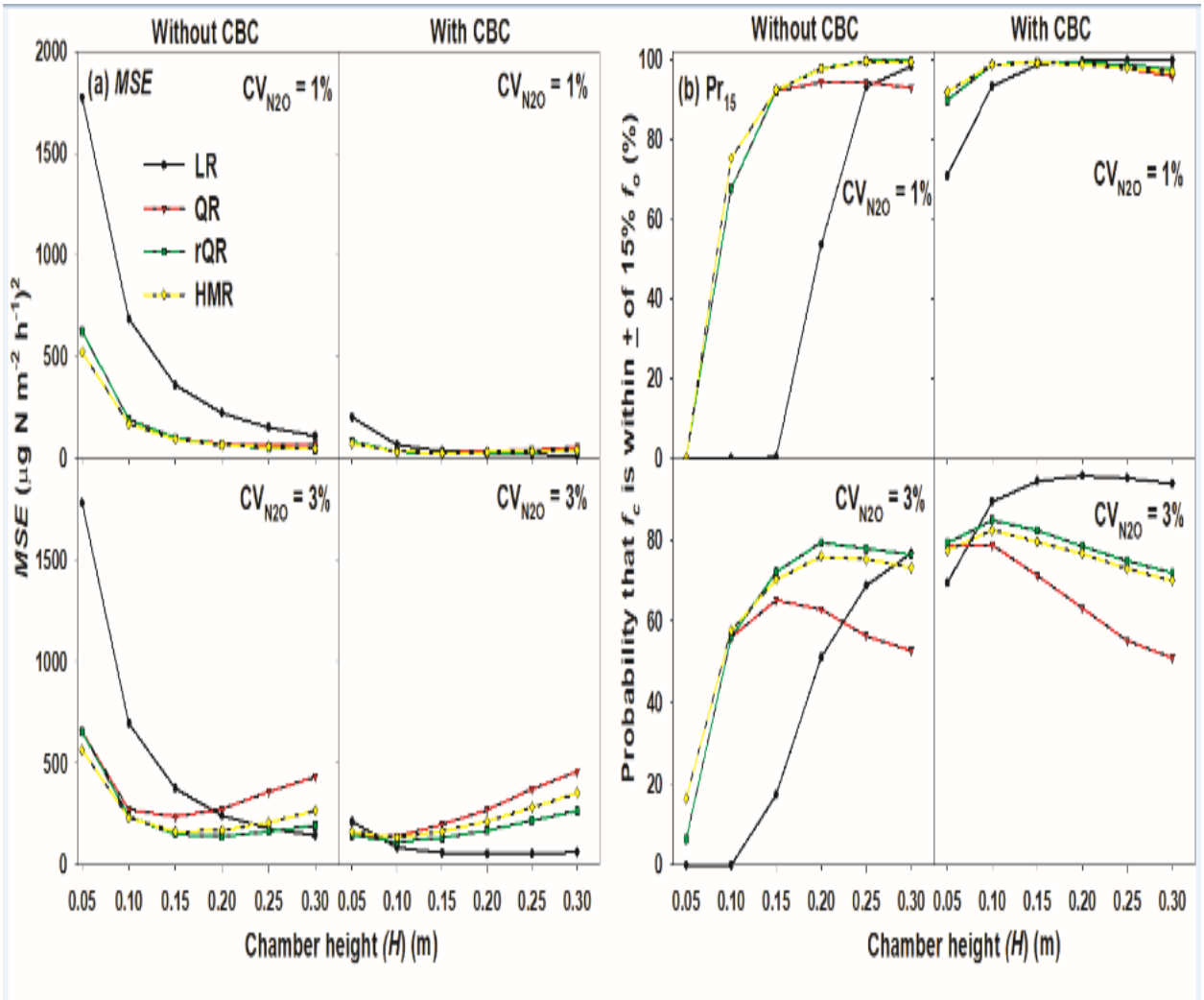

Figure 4. Effects of chamber deployment period (*DP*) and N_2O measurement precision (CV_{N2O}) on (a) mean square error (*MSE*) and (b) probability (Pr_{15}) that the calculated flux (f_c) is within ±15% of the true flux (f_o) for fluxes calculated using linear regression (LR), quadratic regression (QR), restricted quadratic regression (rQR) and HMR with and without chamber bias correction (CBC). The analysis assumed f_o = 100 µg N m^{-2} h^{-1}, chamber height (*H*) = 0.25 m, soil property measurement precision (CV_{SP}) = 10% and soil properties as assumed in Fig. 3. Each *MSE* and Pr_{15} value was determined from a distribution of fluxes calculated from 10,000 Monte Carlo simulated chamber data sets.

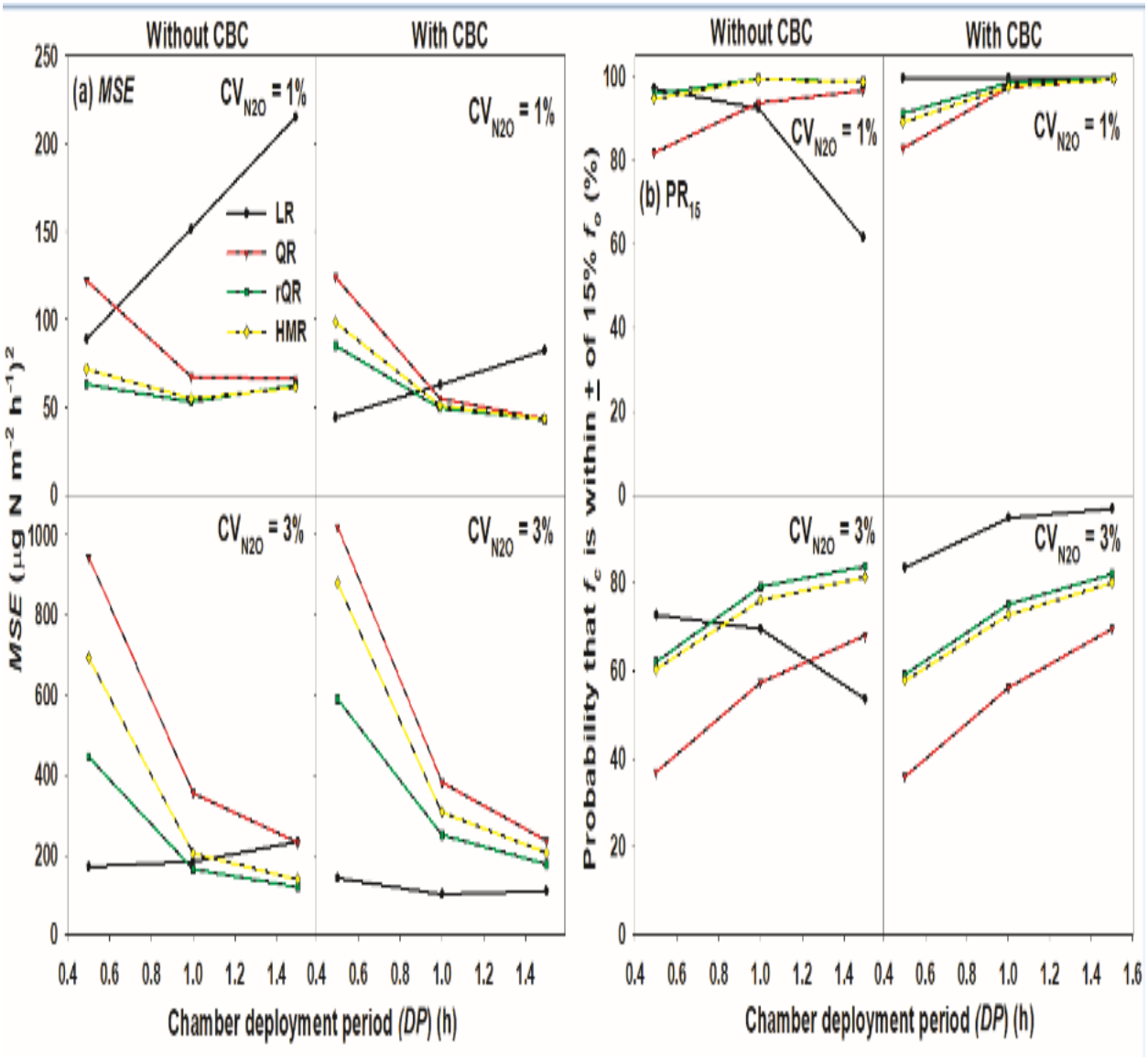

Figure 5. Effects of (a) the true flux (f_o) and (b) soil water content on mean square error (*MSE*) for fluxes calculated using linear regression (LR), quadratic regression (QR), restricted quadratic regression (rQR) and HMR with and without chamber bias correction (CBC). The analysis assumed chamber height (H) = 0.25 m, chamber deployment period (DP) = 1.0 h, N_2O measurement precision (CV_{N2O}) = 3% and soil properties as assumed in Fig. 3-4. Each plotted value was determined from a distribution of fluxes calculated from 10,000 Monte Carlo simulated chamber data sets.

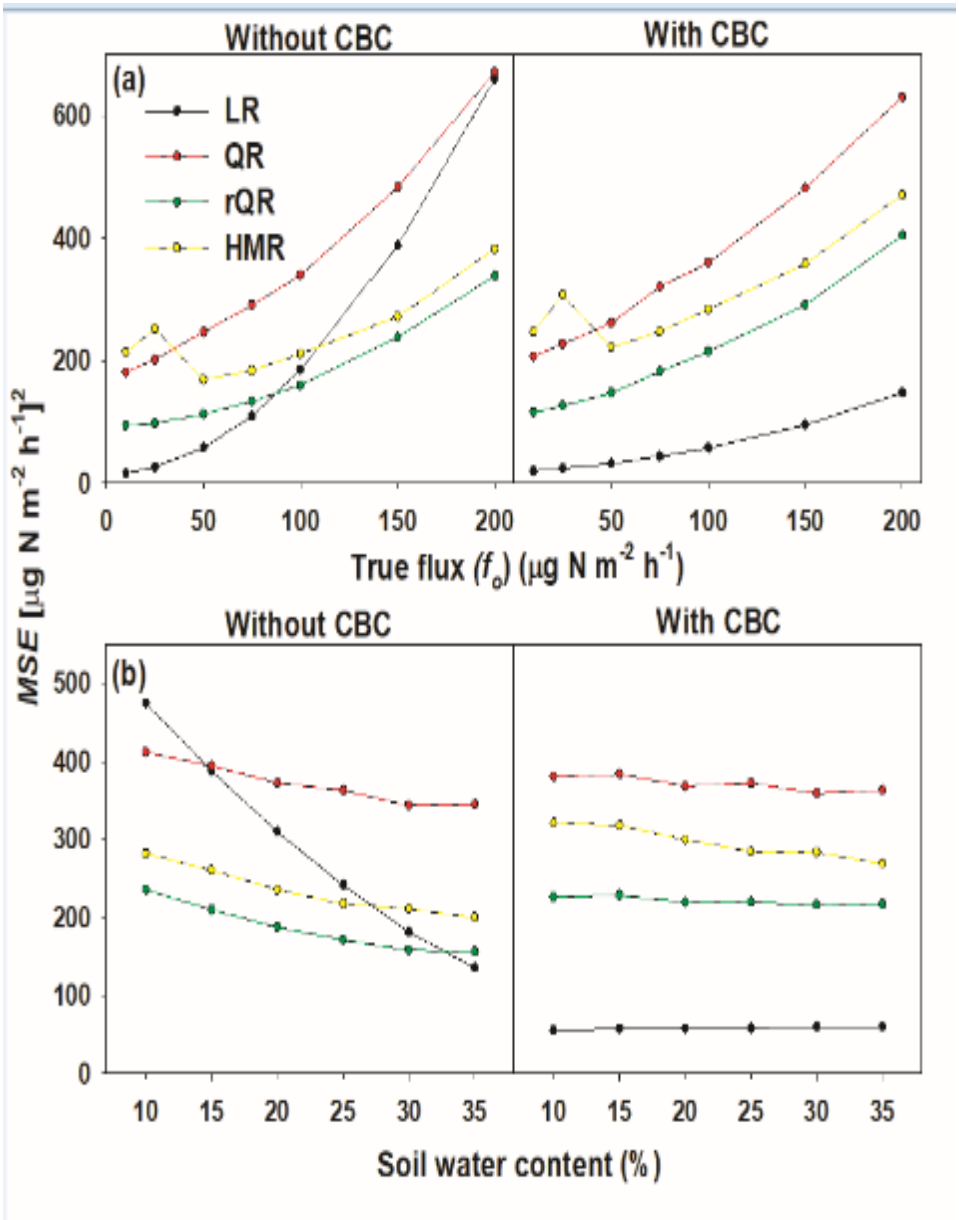

Figure 6. Relationship between the bias (b_{LR}) of fluxes calculated using the linear regression (LR) relative to fluxes calculated using restricted quadratic regression (rQR) or HMR and LR coefficient of determination (r^2) for chamber height (H) = (a) 0.05 m and (b) 0.25 m. Both examples assume chamber deployment period (DP) = 1 h, N_2O measurement precision (CV_{N2O}) = 3%, true flux (f_o) =100 µg N m^{-2} h^{-1} and soil properties as assumed in Fig. 3-5. Results were generated from a distribution of fluxes calculated from 10,000 Monte Carlo simulated chamber data sets where each symbol represents b_{LR} for chamber data sets having r^2 within (a) ±0.004 and (b) ±0.0015 of the plotted r^2 value.